PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

WIPO OMPI

FD

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/27, 5/10, C07K 13/00, 3/28

A2

(11) International Publication Number: WO 94/16075

(43) International Publication Date: 21 July 1994 (21.07.94)

(21) International Application Number: PCT/US94/00658

(22) International Filing Date: 12 January 1994 (12.01.94)

(30) Priority Data:
08/004,141 13 January 1993 (13.01.93) US

(71) Applicant: GENETICS INSTITUTE, INC. [US/US]; Legal Affairs, 87 Cambridge, MA 02140 (US).

(72) Inventors: DORNER, Andrew; 69 Paul Revere Road, Lexington, MA 02173 (US). FRITSCH, Edward, F.; 15 Mildred Circle, Concord, MA 01742 (US). STEININGER, Robert; 100 Reed Street, Cambridge, MA 02140 (US). BUSH, Lawrence; 204 Chapman Road, Tewksbury, MA 01876 (US).

(74) Agent: LAZAR, Steven, R.; Genetics Institute, Inc., Legal Affairs, 87 CambridgePark Drive, Cambridge, MA 02140 (US).

(81) Designated States: AU, BG, BR, CZ, FI, HU, JP, KP, KR, KZ, MG, NO, NZ, PL, RO, RU, SD, SK, UA, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PROCESS FOR PRODUCING M-CSF 223

(57) Abstract

DNA molecules which encode a 255 amino acid polypeptide, comprising the 32 amino acid M-CSF leader sequence and the 223 amino acid mature M-CSF polypeptide, and lack the coding sequence for the carboxy-terminal propeptide sequence present in wild-type M-CSF DNA, result in highly efficient production of M-CSF, which displays similar post-translational modifications as wild-type M-CSF and is unaffected in biological properties.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

PROCESS FOR PRODUCING M-CSF 223

FIELD OF THE INVENTION

The present invention relates to improved processes for producing macrophage-colony stimulating factor (M-CSF or MCSF). More specifically, this invention relates to a method employing a novel DNA sequence encoding directly for a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide, and a recombinant DNA molecule encoding a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide.

BACKGROUND

Colony-stimulating factors are known to be required for the proliferation and differentiation of hematopoietic cells, such as erythrocytes, granulocytes, macrophages, eosinophils, platelets and lymphocytes. MCSF is specific to the formation of macrophages. It is expected that M-CSF would be useful in the treatment of leukocytopenia resulting from chemotherapy or leukocyte levels after bone marrow transplantation. Wild-type M-CSF is expressed in Chinese hamster ovary (CHO) cells as a 522 amino acid membrane anchored precursor protein which is proteolytically processed at amino acid 223 to generate the 45 kDa amino terminal-derived soluble mature form (Figure 1). M-CSF for therapeutic use has been purified as a 90 kDa homodimer whose carboxy-terminus is amino acid 223. In addition, CHO cells expressing the wild-type gene secrete significant amounts of a heterogenous high molecular weight species whose carboxy-terminal sequences extend beyond the residue 223 cleavage site, resulting in high molecular weight heterogeneity of the recombinant protein. Within these extended sequences, glycosaminoglycan addition occurs which contributes to high molecular weight heterogeneity.

The production of human M-CSF (rhM-CSF) by recombinant means has been accomplished using DNA sequences encoding a polypeptide of 554 amino acids. Clark et al., United States Patents 4,868,119 and 4,879,227. This polypeptide is cleaved by mammalian cells to form the 223 amino acid mature M-CSF polypeptide.

Takaku et al., European Patent application (EP) 0 276 551 and United States Patent 5,114,710 disclose the use of 189, 214 and 238 amino acid polypeptide sequences encoding M-CSF to treat thrombocytopenia. However, Takaku et al. do not disclose or suggest the preparation of a 223 amino acid mature M-CSF polypeptide, nor do they disclose

recombinant means or DNA nucleotide sequences encoding a 223 amino acid mature M-CSF molecule. Rather, Takaku et al. disclose methods of purifying M-CSF from human urine. See, EP 0 276 551 at pages 7 to 18.

Kawashima et al., EP 0385 385 discloses that the subunit protein of the homodimer of rhM-CSF has 223 amino acid residues. However, Kawashima et al. disclose methods of preparing 238 amino acid M-CSF polypeptides purified from human urine and does not disclose a nucleotide sequence encoding a 223 amino acid M-CSF polypeptide.

Toyo Jozo Co., UK Patent Application (GB) 2 249 100 A discloses a recombinant DNA vector capable of expressing a 179 amino acid M-CSF polypeptide.

Kawasaki et al., Science, 230:291-296 (1985) disclose a recombinant DNA sequence encoding a 256 amino acid M-CSF polypeptide, which includes a 32 amino acid leader sequence. This 224 amino acid mature protein (after removal of the leader sequence) is structurally distinct from the 223 amino acid form disclosed here in that amino acids 150 through 224 are different from amino acids 150 to 223 of the M-CSF polypeptide of the present invention.

SUMMARY OF THE INVENTION

To increase the productivity of cell lines for the 90 kDa homodimer, we have developed an expression vector designated pMemcMAT which is designed to produce only the 255 amino acid polypeptide comprisng the 32 amino acid M-CSF leader and the 223 amino acid mature form of M-CSF (Figure 3). Surprisingly, the 255 amino acid form of M-CSF is properly processed when expressed in CHO cells, despite the fact that the polypeptide is not synthesized as part of a membrane-bound precursor protein due to the deletion of the carboxy-terminal sequences. Importantly, such cell lines do not produce the high molecular weight proteoglycan form of M-CSF.

Accordingly, it is one object of the present invention to provide recombinant DNA vectors capable of directly expressing a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide.

It is another object of the present invention to provide improved mammalian cell lines capable of directly expressing a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide.

It is yet another object of the present invention to provide improved processes for the production of M-CSF by culturing transformed mammalian cell lines which are capable of directly expressing a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide.

It is one feature of the present invention that cell lines transformed with DNA directly encoding a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide are used.

It is one advantage of the present invention that M-CSF is produced with improved yield and process efficiency.

It is another advantage of the present invention that the M-CSF produced is post-translationally modified (e.g., N-linked and O-linked glycoylation) similarly to rhM-CSF expressed from the full-length cDNA sequence.

It is another advantage of the present invention that the high molecular weight heterogeneity normally occurring with the production of rhM-CSF in wild-type cell lines is largely or wholly eliminated.

Other objects, features and advantages of the present invention will become more readily apparent from the following description.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: The amino acid sequence of wild type M-CSF is shown. Asterisks and arrows (^) indicate the position of the inserted termination codons and EcoRI digestion site, respectively. The potential chondroitin sulfate addition site in the propeptide region is underlined.

Figure 2: Expression vector pMemc for wild-type M-CSF. pMemc contains the dihydrofolate reductase (DHFR) gene, which confers resistance to methotrexate. Transcription of a dicistronic mRNA is directed by the adenovirus major late promoter (MLP). The 5' end of the mRNA contains the adenovirus tripartite leader sequence (TPL)

followed by the full length M-CSF coding sequence (MCSF). Translation of the 3' proximal selectable marker gene is under control of the Encephalomyelocarditis virus leader sequence (EMC-L). The vector also contains the SV40 origin of replication and enhancer (SV40 ori), adenovirus VAI gene (VA) and the SV40 polyadenylation site (SV40-pA). Selected restriction digestion sites are indicated.

Figure 3: Expression vector for MCSF-223. Termination codons and an EcoRI restriction site were inserted immediately following the coding sequences for the carboxy terminus of the 45 kDa monomer at amino acid residue 223 by site-directed DNA mediated mutagenesis of pMemc. EcoRI digestion was performed to delete the propeptide coding sequences to generate pMemcMAT. The same regulatory elements are present on pMemcMAT as described for pMemc in Figure 2.

DESCRIPTION OF THE SEQUENCES

SEQUENCE ID NO. 1 is the nucleotide sequence encoding the 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide.

SEQUENCE ID NO. 2 is the amino acid sequence of the 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the MCSF-223 polypeptide.

SEQUENCE ID NO. 3 is the amino acid sequence of the 186 amino acid polypeptide comprising the DHFR coding sequence which confers resistance to methotrexate and may be used as a selectable marker.

SEQUENCE ID NO. 4 is the amino acid sequence of the wild-type M-CSF polypeptide, including the leader sequence and 522 amino acid.

DETAILED DESCRIPTION OF THE INVENTION

In order to accomplish the above objectives, the inventors followed a program first of generating new cell lines. This program comprised construction of vectors for the production of the 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide; selection, amplification and cloning of

transformed cells; and measuring the productivity of the cells thus obtained. The initial cell lines and proteins produced thereby were then characterized.

Second, high productivity cell lines were characterized and adapted to cell culture conditions. Next, cellular and volumetric productivity were measured, followed by identification of growth characteristics.

Third, the 223 amino acid mature M-CSF protein, as directly expressed by the DNA sequence encoding the 255 amino acid polypeptide, comprising the 32 amino acid M-CSF leader and the 223 amino acid mature M-CSF polypeptide, was characterized and compared to the protein produced by expression of the full length (wild type) M-CSF polypeptide following intracellular proteolytic processing. N-terminal and C-terminal amino acid sequences were analyzed. The protein was analyzed by Achrobacter protease K peptide mapping. The N-glycan fingerprint was determined. Chromatographic analyses, including reversed phase and size exclusion HPLC, were performed. Electrophoretic analyses, including polyacrylamide gel electrophoresis and isoelectric focusing, were performed.

Fourth, the biological activity of the M-CSF produced by the DNA sequences of the present invention was measured and compared to that of M-CSF produced using wild type cell lines. <u>In vitro</u> activity was measured in the 32Dcfms assay and murine bone marrow assay. Half-life, clearance rate, and volume of distribution of M-CSF polypeptide in rats was measured. Finally, incidence of monocytosis and thrombocytopenia in the rat was observed.

Accordingly, the present invention comprises isolated and purified DNA molecules which consist essentially of the DNA sequence of Sequence ID No. 1. The DNA molecules of the present invention encode the production of a 255 amino acid polypeptide, comprising the 32 amino acid M-CSF leader and a 223 amino acid polypeptide having M-CSF activity and appropriate post-translational modifications, said polypeptide lacking the carboxy-terminal propeptide sequence, including a transmembrane domain, present in wild-type recombinant M-CSF precursor polypeptide before intracellular processing when produced by

mammalian cells. In a preferred embodiment, the DNA molecule of the present invention comprises the vector pMemcMAT.

The present invention also comprises isolated and purified DNA molecules consisting essentially of a DNA sequence encoding the polypeptide comprised of the amino acid sequence of Sequence ID No. 2. Because of the degeneracy of the genetic code, it will be recognized that numerous changes to the DNA sequence of Sequence ID No. 1 can be made without altering the amino acid sequence of the polypeptide encoded by said DNA sequence. See, for example, Lehninger, Biochemistry, 2d ed., pp.959-963 (Worth Publishers, New York 1975). The DNA molecules of the present invention encode the production of a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader sequence and a 223 amino acid polypeptide having M-CSF activity, said polypeptide lacking the carboxy-terminal propeptide sequence, including a transmembrane domain present in wild-type recombinant M-CSF precursor polypeptide before intracellular processing when produced by mammalian cells.

The present invention also comprises methods of producing MCSF-223 protein comprising culturing in a suitable culture medium cells transformed with a DNA molecule comprising the DNA sequence of Sequence ID No. 1 and isolating and purifying said protein from said culture medium. In a preferred embodiment, the method of the present invention comprises culturing in a suitable culture medium cells transformed with pMemcMAT and isolating and purifying said protein from said culture medium. Suitable culture media are known in the art and depend upon the species of the host cells to be used to produce the MCSF-223 protein. Where the host cells to be used are Chinese Hamster Ovary (CHO)

cells, for example, suitable culture media include those described in Hamilton and Ham, In Vitro, 13: 537-547 (1977) and Mendiaz et al., In Vitro Cellular and Developmental Biology, 22: 66-74 (1986).

In another embodiment, the present invention comprises mammalian cell lines capable of producing MCSF-223 protein, said mammalian cell line comprising a DNA sequence directly encoding a 255 amino acid polypeptide comprising the 32 amino acid M-CSF leader sequence and the 223 amino acid mature M-CSF polypeptide, said polypeptide lacking the carboxy-terminal propeptide sequence present in wild-type recombinant M-CSF precursor polypeptide before intracellular processing when produced by mammalian cells.

The present invention further comprises other improvements to the process for producing M-CSF polypeptide. These improvements include the use of ultrafiltration/diafiltration steps, the use of alternate resins, and the development of the manufacturing process.

Definitions

As used in the present application, the term "M-CSF activity" refers to at least one of the "CSF-1-like biological properties" as defined in United States Patent 4,868,119, the text of which is hereby incorporated herein by reference.

As used in the present application, the term "MCSF-223", "MCSF-223 protein" or "MCSF-223 polypeptide" refers to an amino acid sequence with M-CSF activity, which is 223 amino acids in length, but which is free of the carboxy-terminal propeptide sequence normally present in wild-type M-CSF before intracellular processing when produced by

mammalian cells. The MCSF-223 protein of the present invention is produced through recombinant means from MCSF-223 DNA, without the carboxy-terminal propeptide.

As used in the present application, the terms "MCSF-223 vector" and "MCSF-223 DNA vector" refer to recombinant DNA vectors capable of directly expressing a 223 amino acid mature M-CSF polypeptide. Such an MCSF-223 vector will comprise a coding DNA sequence of approximately 765 nucleotides in length, such that it will encode the 223 amino acid mature M-CSF polypeptide and include DNA encoding the 32 amino acid leader peptide which is considered to encompass amino acids -32 to -1 of Figure 1. It will be understood in the art, such vectors may further comprise DNA sequences not directly encoding the M-CSF mature protein, such as regulatory elements, e.g., promoters, leaders, or enhancers, and/or DNA sequences unrelated to M-CSF, e.g., signal peptides and selectable marker genes. However, the MCSF-223 vectors are free from DNA encoding the carboxy-terminal propeptide normally present in wild-type M-CSF precursor protein.

As used in the present application, the term "MCSF-223 cell line" refers to mammalian cell lines capable of directly expressing an MCSF-223 polypeptide. The MCSF-223 cell lines of the present invention are transformed with MCSF-223 vector DNA, so that they directly encode the production of MCSF-223.

As used in the present application, the terms "wild-type M-CSF", "wild-type M-CSF cell line" and "wild type M-CSF DNA vector" refer to DNA, vectors and cell lines which produce an M-CSF polypeptide significantly longer than 223 amino acids in length. The wild-type M-CSF precursor protein is 522 amino acids in length and contains a carboxy-terminal propeptide sequence which is normally cleaved during intracellular processing by

mammalian cells to generate the mature form of 223 amino acids. However, a significant proportion is cleaved at sites beyond 223 to generate high molecular weight M-CSF proteoglycan species.

Examples

Example 1. Construction of M-CSF expression vectors.

A 1790 bp human M-CSF cDNA fragment derived from the M-CSF expression plasmid p3AMCSF-R1 was inserted into the expression vector pED to generate the vector pMemc. (Figure 1). pMemc contains the dihydrofolate reductase (DHFR) gene which confers resistance to methotrexate (MTX). Transcription to generate a dicistronic mRNA is directed by the adenovirus major late promoter in combination with the SV40 enhancer. The 5' end of the mRNA encodes M-CSF while the 3' end encodes the selectable and amplifiable marker DHFR. Translation of the 3' proximal gene is under the control of the encephalomyelocarditis virus (EMCV) leader sequence. The presence of this sequence between the M-CSF coding sequence and the selectable marker allows efficient internal entry of ribosomes to initiate protein synthesis of the selectable marker.

To increase the productivity of cell lines for the 90 kDa homodimer, we have developed an expression vector designated pMemcMAT to produce only the 255 amino acid polypeptide, comprising the 32 amino acid M-CSF and the mature 223 amino acid form of M-CSF (Figure 2). This was accomplished by site directed DNA mediated mutagenesis of pMemc. Protein synthesis termination codons and an EcoRI restriction site were inserted immediately following the coding sequence for amino acid residue 223. EcoRI digestion of the vector resulted in removal of the 3' terminal sequences encoding the extended carboxy-

terminal propeptide region. Removal of these sequences precludes potential readthrough translation of the termination codons to generate extended protein products. pMemcMAT directs expression of the mature form of M-CSF of 223 amino acid residues and confers resistance to MTX. pMemcMAT was deposited with the ATCC on December 11, 1992 and has been accorded ATCC accession number 75378. **Example 2. Development of CHO cell lines expressing M-CSF.**

A. Development and characterization of MCSF-223 producing cell lines.

pMemcMAT was introduced into CHO DUKX cells by lipofection and cells were subsequently selected for growth in the absence of nucleosides and the presence of 0.02 uM or 0.1 uM MTX. MTX resistant cells were picked into 16 independent pools and subjected to further rounds of selection and amplification in increasing concentrations of MTX. M-CSF production was monitored primarily using an ELISA assay developed at Genetics Institute which detects total M-CSF antigen. The highest producing cells were identified and subcloned by limiting dilution to produce clonal cell lines for further development. (Table I). The cellular productivity of these cell lines was quantitated by both ELISA assay and size exclusion chromatography (SEC) to monitor the levels of total M-CSF antigen and 90 kDa dimer, respectively. Cellular productivity ranged from about 30 to about 70 pg/cell/24h of total secreted antigen for individual cell lines. This represents an approximate 10-25 fold increase in productivity compared to previous production using wild type expressing cell line. The current wild type cell line produces about 3-5 pg/cell/24h under similar assay conditions.

TABLE I

MCSF-223 CHO Cell Lines			
Cell line	**MTX (uM)**	**Total antigen pg/cell/24hr**	**90 kDa dimer pg/cell/24hr**
Line H1	2	30	28
Line H8	5	40	42
Line H1L	5	70	ND
Line C10	5	30	20
Line D2	2	50	40
Line D1	5	40	ND

Legend: MCSF-223 cell lines were fluid changed into complete alpha medium (total antigen) or alpha defined medium without BSA (90 kDa dimer). Conditioned medium was harvested and cell numbers determined 24 h later. Cellular productivity was calculated by dividing volumetric productivity by cell number. MCSF total antigen levels were determined by ELISA. Levels of 90 kDa dimer were determined by size exclusion chromatography (SEC). Both values are given in pg/cell/24h. **MTX**, methotrexate; **ND**, not determined.

Recent reports have shown that the high molecular weight heterogenous M-CSF species contains a proteoglycan with carboxy terminal sequences which extend beyond the 223 cleavage site. Two potential glycosaminoglycan sites are carboxy-terminal to the cleavage site at amino acid 223 (see Figure 1). Thus, expression of the MCSF-223 polypeptide form should result in the absence of the high molecular weight proteoglycan species since the carboxy-terminal sequences including the glycosaminoglycan site are absent.

Analysis of unlabeled conditioned medium by SEC confirms lesser quantitites of high molecular weight material secreted from MCSF-223 producing cells compared to wild type M-CSF. Two distinct peaks of high molecular weight material were detected by SEC of medium conditioned by MCSF-223 producing cells. Western blot analysis indicated that the material in one peak is M-CSF, which upon reduction has the electrophoretic mobility of 45 kDa monomer, while the other HMW peak is CHO cell specific protein. This HMW material is presumably different from that HMW material present in the wild-type cell lines. Approximately 89% of the material retained between 10 and 15 minutes consists of 90 kDa homodimer. In sharp contrast, SEC analysis of medium conditioned by wild type M-CSF producing cells detected a much higher proportion of high molecular weight species. This high molecular weight species does not display an electrophoretic mobility of 45 kDa following reduction, indicating it is different from the small amount of high molecular weight material observed for MCSF-223. This represents a significant advantage of the MCSF-223 cell lines, from both a yield and protein purification perspective, over wild-type M-CSF cell lines.

Example 3. Secretion of M-CSF in Bioreactors.

Cell line H1 (Table I) was adapted to serum-free suspension culture. In petri dishes, 90 kDa M-CSF was produced by the resulting cell line at 14.7 pg/cell/day, whereas 90 kDa M-CSF was proudced by a wild-type M-CSF cell line at only 2.0 pg/cell/day. The adapted H1 cell line was also grown in a 250-L bioreactor. In a series of seven 3-day or 4-day batch growth cycles, 90 kDa M-CSF was produced at 10.2 pg/cell/day, whereas in a series of 22

3-day batch growth cycles, 90 kDa M-CSF was produced by the wild type cell line at only 1.6 pg/cell/day. Mean cell densities were as high or higher during the cycles using the H1 cell line than during the cycles using the wild type cell line. Thus, the improvement in cellular productivity (pg/cell/day) leads to an improvement in overall productivity (g/day) cell line H1 has been deposited with the ATCC and has been accorded ATCC designation CRL 11275.

Example 4. Protein Characterization.

Critical to the development of an MCSF-223 producing cell line was comparison of the MCSF-223 to the wild type M-CSF material. This characterization process required purification of M-CSF produced from various MCSF-223-producing cell lines and from wild type cell lines. Material from cells grown in roller bottles was purified by DEAE/Q-Sepharose tandem chromatography followed by Jacalin agarose affinity chromatography. Final purification was done using Sephacryl S-300 chromatography. This purification process confirmed that MCSF-223 cell lines secreted predominantly the 90 kDa dimer species of M-CSF with little high molecular weight material (< 10%) and no propeptide in the conditioned medium. High molecular weight M-CSF and propeptide are the major contaminants of the wild-type M-CSF process. This indicates that purification of MCSF-223 is preferable to wild type M-CSF. Purified material was then characterized by a variety of methods as outlined below.

Amino-terminal and carboxy-terminal sequence analysis of MCSF-223 from a single cell line indicated that the predominant amino-terminal amino acid sequence corresponded

to the mature terminus beginning with EEVSE and that the predominant carboxy-terminal amino acid sequence corresponded to the correct mature terminus of RPPR. This important result indicates that termination of the protein at amino acid residue 223 does not result in aberrant proteolytic processing to generate truncated carboxy-terminal species.

Peptide mapping of Achrobacter protease K digested MCSF-223 from line H1 compared to wild type M-CSF from current wild type cell line showed similar but not identical profiles. All peptides present in the MCSF-223 chromatogram were also present in the wild type chromatogram and no peptides were absent inclusive of both the amino-terminus and carboxy-terminus peptides. Since the amino-terminus and carboxy-terminus of MCSF-223 correspond to those of wild type based upon amino acid sequence analysis and peptide maps were similar, no major structural differences appear to exist between MCSF-223 and wild-type M-CSF protein. It was observed that MCSF-223 displayed a distribution of peptides K12K13 and K14 different from that of the wild-type M-CSF. These peptides from wild-type M-CSF are known to reflect variability in the amount of O-linked glycosylation and sialation. The profile from MCSF-223 shows increased amounts of presumably more highly O-glycosylated and/or sialated peptides which are naturally present at lower levels in wild-type M-CSF. This analysis suggests that MCSF-223 contains increased but not abnormal O-linked glycosylation and/or sialation compared to wild-type M-CSF.

Comparative analysis of N-linked glycosylation of MCSF-223 from several cell lines compared to wild-type M-CSF from Mα3-18 cell line was performed using high pH anion exchange chromatography (HPAEC) of the N-glycans liberated from the protein by PNGase

digestion. Chromatograms of a representative MCSF-223 polypeptide derived from MCSF-223 line H1 cells and wild-type M-CSF derived from current wild type cells demonstrate that the spectrum of N-linked glycans is quite similar, but not identical, between MCSF-223 and wild-type forms of M-CSF. Specifically, there exist two areas of the N-linked fingerprint which display slight differences between MCSF-223 and wild-type M-CSF. MCSF-223 exhibits lower levels of monosialyted species eluting at approximately 20 minutes in the profile. Slight variation in peak ratios exists in the trisialyted species eluting at approximately 40 minutes in the profile. The exact structural basis for these differences is not known, but it is believed that differences in terminal sialyation are responsible. Preliminary data obtained by Warren assay to determine sialic acid content indicates that MCSF-223 contains more sialic acid than wild-type. These characterization studies indicate that carbohydrate addition and processing of MCSF-223 is not significantly altered compared to wild-type. MCSF-223 displays increased O-linked glycosylation and sialation. However, this appears to be the result of increased levels of carbohydrate moieties naturally present in the wild type M-CSF population rather than the appearance of species unique to MCSF-223.

Analysis of MCSF-223 by SEC indicated that the amounts of high molecular weight (HMW) species present in conditioned medium was, as expected, lower than in wild type M-CSF producing cells. However, a small amount of high molecular weight material was present in the medium. The reduction of the HMW species produced by the MCSF-223 cell lines to 45 kDa MW differentiates it from the HMW species produced by wild-type cell lines. Wild type HMW material upon reduction exhibits a mw of 150-200 kDa due to the presence of the glycosaminoglycan and the extended carboxy-terminal sequences.

Example 5. Biological Activity.

The above data demonstrate that the MCSF-223 and wild-type M-CSF mature forms are similar but not identical. An experimental program was designed to compare the biological activities of the proteins. The results of the program, as described below, demonstrate that the two materials are extremely similar, if not identical, in the biological properties tested. The in vitro specific activity of purified M-CSF from MCSF-223 cell lines and wild type cell lines was determined in the 32Dcfms assay (Table II). This assay determines the activity of M-CSF by measuring its ability to stimulate the proliferation of 32D murine myeloid cells which stably express the human M-CSF receptor (c-fms). Material was assayed in a full range dose response curve with a purified M-CSF standard as reference. The slopes of the curves for the various preparations were similar allowing the calculation of specific activity using the ED_{50} value. The in vitro specific activity of the purified MCSF-223 was similar to that of wild type M-CSF and within the variation of the assay. The ED_{50} of the standard has been defined as 19.9 ± 7.9 ng, the specific activity has been defined as $0.6 \pm 0.23 \times 10^5$ U/mg, and the slope as 1.2 ± 0.3. In addition, purified M-CSF from four MCSF-223 cell lines was tested in the mouse bone marrow assay. This assay determines the ability of M-CSF to stimulate proliferation of hematopoietic progenitor cells derived from murine bone marrow. All MCSF-223 samples were shown to have biological activity of $> 0.5 \times 10^6$ U/mg which is similar to that obtained for wild type M-CSF.

TABLE II

M-CSF Bioassay Results for MCSF-223 and Wild-type M-CSF cell lines			
Cell line	Slope	ED_{50}ng	S.A. U/mg
MCSF-223 Line H1	1.28	27.7	$0.36x10^5$
MCSF-223 Line D2	1.32	17.6	$0.57x10^5$
Wild-type M-CSF	1.31	27.3	$0.37x10^5$
Standard (Wild-type)	1.30	19.3	$0.52x10^5$

Legend: S.A., specific activity.

Example 6. Evaluation of Pharmacokinetics and Biologic Activity of MCSF-223.

The goal of these evaluations was to compare the pharmacokinetic and biologic activity of rhM-CSF produced by MCSF-223 cell lines to rhM-CSF produced by wild type M-CSF cell line. A two-tiered approach was used for this evaluation. The first step was a rat pharmacokinetic profile of rhM-CSF produced by four different MCSF-223 cell lines compared to control rhM-CSF. Two of the MCSF-223 cell line rhM-CSF samples used in the pharmacokinetic evaluation were then selected for a secondary evaluation of biologic activity in rats following intravenous and subcutaneous administration.

Pharmocokinetic Experimental Design.

Twenty female Sprague-Dawley rats were used for this evaluation. The rats were randomly divided into 5 groups of four rats each. The test groups received an intravenous

bolus injection of iodinated MCSF-223 protein produced by MCSF-223 cell lines. A control group received iodinated rhM-CSF produced by wild-type M-CSF. Rats were anesthetized by an IM injection of a mixture of Ketamine and Xylazine and dosed by tail vein injection with a mixture of ^{125}I rhM-CSF and unlabelled rhM-CSF at a total dose of 50 ug/kg. Blood was taken at 0.5,2,5,10,15,45,90,180 and 360 minutes post dosing and was allowed to clot at room temperature for 10 minutes. The samples were kept on ice for ten minutes and then centrifuged to separate the serum. Aliquots of serum (50ul) were counted for ^{125}I in a gamma counter. Precipitable counts were determined in 20% TCA.

Serum rhM-CSF level versus time curves were constructed and fit by a mono-exponential equation of the form $Cp=A*e^{-kt}$ by a non-linear regression program. The half-life, volume of distribution and total body clearance for rhM-CSF in each animal were determined based on the slope and intercept of the best fit function.

The serum samples of selected rats (N=2/Group) were also evaluated by SEC-HPLC. Serum (20ul) was injected directly onto a SEC-250 column (Biorad) and ^{125}I was monitored using an on-line gamma detector (Berthold). The height of a single peak that eluted at the same apparent molecular weight as rhM-CSF was measured for each serum sample. The percent ^{125}I rhM-CSF remaining was determined using the 0.5 minute peak height as 100%. Pharmacokinetic analysis was done as described above.

The serum half-life of precipitable radioactivity for rhM-CSF from MCSF-223 cell lines ranged from 54 minutes to 83 minutes. The half-life and clearance estimates of the MCSF-223 cell line variants were not statistically different from control values (Students T-test). The pharmacokinetic parameters of half-life, total body clearance, and volume of

distribution are summarized in Table III. The volume of distribution approximated serum volume and was similar for all rhM-CSF variants evaluated. The half-life values determined by SEC-HPLC were very similar to those obtained from precipitable serum radioactivity. Serum radioactivity eluted from the column as a single peak at the same retention time as rhM-CSF, suggesting that the precipitable counts in the serum represent intact rhM-CSF.

TABLE III

PHARMACOKINETIC SUMMARY OF MCSF-223 CELL LINES IN SPRAGUE DAWLEY RAT MEAN +/- SD				
CELL LINE	1/2-LIFE (MIN)	C_{MAX}	TOTAL CLEARANCE (mL/MIN/KG)	VOLUME OF DISTRIBUTION (mL)
Line D2	54 4	877 83	0.75 0.10	19 2
Line C10	53 10	975 57	0.69 0.12	17 1
Line H1	69 7	916 71	0.55 0.10	18 1
Line H8	83 11	909 53	0.47 0.06	17 1
Control (wild-type)	70 15	873 32	0.60 0.16	18 1

The lower number in each box represents the standard deviation

Example 7. Biologic Effects of MCSF-223 form rhM-CSF

Forty-eight female Sprague Dawley rats were divided randomly into eight groups of six animals each. Four groups received test article or vehicle control as a single subcutaneous injection of 2.5 mg/Kg. Four groups received test article or vehicle control as daily intravenous injections of 500 ug/kg/day for 5 consecutive days. The groups received MCSF-223 form of rhM-CSF, control rhM-CSF produced by wild-type M-CSF cell lines or vehicle control. The animals were anesthetized with isoflurane inhalant anesthetic and bled by cardiac puncture on days 0,2,4,6,8 and 15 onto EDTA. Hematological parameters, measured on a Baker 9000 hematology analyzer, included white blood cell count, red blood

cell count, platelet count, hemoglobin, hematocrit, and Weintrobe's constants. Differential cell counts (100 cells) were performed on Wrights-Geisma stained peripheral blood smears.

A decrease in platelet count has been the most reproducible observation following the administration of rhM-CSF to rats. The subcutaneous groups showed a maximum decrease on day 2 after dosing of between 30 and 40 percent below baseline. The kinetics of platelet decrease and recovery were similar for all subcutaneous rhM-CSF receiving groups (MCSF-223 cell lines and wild-type rhM-CSF cell lines). The intravenous groups consistently reached a low platelet count on day 4 at about 50 percent below baseline. The platelet decrease and rebound was similar for animals receiving MCSF-223 cell line rhM-CSF and control (wild-type cell line) rhM-CSF intravenously. No significant platelet effects were noted in the vehicle receiving groups.

A peripheral monocytosis is observed following the subcutaneous or intravenous administration of rhM-CSF to rats, but historically, this response has been more variable than the very consistent thrombocytopenia. Both dosing routes resulted in a variable peripheral monocytosis which reached a maximum on day 2 of the evaluation. No consistent changes were noted in the vehicle receiving group. MCSF-223 produced by the MCSF-223 cell line H1 showed the largest increase in total peripheral monocytes for both dosing routes.

Following intravenous administration to rats, the pharmacokinetic profile of rhmcsf produced by MCSF-223 cell lines were not statistically different from rhmcsf produced by wild-type M-CSF cell line. MCSF-223 produced by the MCSF-223 cell lines H1 and D2 produced platelet decreases following intravenous and subcutaneous administration to rats that were indistinguishable from that produced by rhM-CSF produced by wild-type M-CSF cell

ultrafiltration/diafiltration step has been incorporated prior to the initial QAE-Sephrose capture step. The diafiltration removes low molecular weight components and substantially enhances the capacity of the QAE-Sepharose column, allowing much smaller columns to be used. Additionally, ceramic Hydroxyapatite has been substitute for Hydroxyapatite Ultragel. The ceramic matrix has significantly higher capacity as well as more desirable handling properties. WIth the changes described above, the purificaiton process for purification of the M-CSF from the 223 cell line now consists of ultrafiltration/diafiltration; then sequential chromatography on QAE-Sephrose, ceramic Hydroxyapatite and Phenyl-Toyopearl; and finally diafiltration to yield bulk drug substance in the appropriate buffer. This process provides a robust purification scheme, readily able to handle the large amounts of M-CSF prodluced by the 223 cell line and able consistently to yield a high purity product.

Example 9. 250 L Bioreactor/Purification Scale-up Runs

A 250 L bioreactor was used to produce six successful harvests from M-CSF cell lines. Each run has resulted in improvements with the % recovery approaching 95%. Chromatographic purification was used to produce 1.3 runs. The major quantitative contaminant detected is the HMW-MCSF, at less than about 1%, a significant improvement over prior processes.

__Summary__

Extensive characterization studies of have shown that no significant differences exist in biological characteristics and activity purified MCSF-223 protein compared to wild type M-CSF. The absence of high molecular weight proteoglycan species of M-CSF produced from MCSF-223 cell lines represents a significant advantage from both yield and protein

purification perspectives. We have shown that, using the MCSF-223 vector, volume productivity in production of M-CSF protein has been increased at least approximately 10-fold.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Dorner, Andrew
Fritsch, Edward
Steininger, Robert
Bush, Lawrence

(ii) TITLE OF INVENTION: MCSF-223 Amino Acid Process

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
(B) STREET: 87 CambridgePark Drive
(C) CITY: Cambridge
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02140-2387

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 08/004,141
(B) FILING DATE: 13-JAN-1993
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Lazar, Steven R.
(B) REGISTRATION NUMBER: 32,618
(C) REFERENCE/DOCKET NUMBER: GI-5210

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 617-876-1170
(B) TELEFAX: 617-876-5851

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Dorner, Andrew
Fritsch, Edward
Steininger, Robert
Bush, Lawrence

(ii) TITLE OF INVENTION: MCSF-223 Amino Acid Process

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
(B) STREET: 87 CambridgePark Drive
(C) CITY: Cambridge
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02140-2387

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: 08/004,141
(B) FILING DATE: 13-JAN-1993
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Lazar, Steven R.
(B) REGISTRATION NUMBER: 32,618
(C) REFERENCE/DOCKET NUMBER: GI-5210

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 617-876-1170
(B) TELEFAX: 617-876-5851

Ala Lys Gln Arg Pro Pro Arg
250 255

TCCCTCCCCC CCCCCTAACG TTACTGGCCG AAGCCGCTTG GAATAAGGCC GGTGTGCGTT 2086

TGTCTATATG TTATTTTCCA CCATATTGCC GTCTTTTGGC AATGTGAGGG CCCGGAAACC 2146

TGGCCCTGTC TTCTTGACGA GCATTCCTAG GGGTCTTTCC CCTCTCGCCA AAGGAATGCA 2206

AGGTCTGTTG AATGTCGTGA AGGAAGCAGT TCCTCTGGAA GCTTCTTGAA GACAAACAAC 2266

GTCTGTAGCG ACCCTTTGCA GGCAGCGGAA CCCCCCACCT GGCGACAGGT GCCTCTGCGG 2326

CCAAAAGCCA CGTGTATAAG ATACACCTGC AAAGGCGGCA CAACCCCAGT GCCACGTTGT 2386

GAGTTGGATA GTTGTGGAAA GAGTCAAATG GCTCTCCTCA AGCGTATTCA ACAAGGGGCT 2446

GAAGGATGCC CAGAAGGTAC CCCATTGTAT GGGATCTGAT CTGGGGCCTC GGTGCACATG 2506

CTTTACATGT GTTTAGTCGA GGTTAAAAAA CGTCTAGGCC CCCCGAACCA CGGGGACGTG 2566

GTTTTCCTTT GAAAAACACG ATTGCTCGAG CCATC ATG GTT CGA CCA TTG AAC 2619
Met Val Arg Pro Leu Asn
1 5

TGC ATC GTC GCC GTG TCC CAA AAT ATG GGG ATT GGC AAG AAC GGA GAC 2667
Cys Ile Val Ala Val Ser Gln Asn Met Gly Ile Gly Lys Asn Gly Asp
10 15 20

CTA CCC TGG CCT CCG CTC AGG AAC GAG TTC AAG TAC TTC CAA AGA ATG 2715
Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe Lys Tyr Phe Gln Arg Met
25 30 35

ACC ACA ACC TCT TCA GTG GAA GGT AAA CAG AAT CTG GTG ATT ATG GGT 2763
Thr Thr Thr Ser Ser Val Glu Gly Lys Gln Asn Leu Val Ile Met Gly
40 45 50

AGG AAA ACC TGG TTC TCC ATT CCT GAG AAG AAT CGA CCT TTA AAG GAC 2811
Arg Lys Thr Trp Phe Ser Ile Pro Glu Lys Asn Arg Pro Leu Lys Asp
55 60 65 70

AGA ATT AAT ATA GTT CTC AGT AGA GAA CTC AAA GAA CCA CCA CGA GGA 2859
Arg Ile Asn Ile Val Leu Ser Arg Glu Leu Lys Glu Pro Pro Arg Gly
75 80 85

GCT CAT TTT CTT GCC AAA AGT TTG GAT GAT GCC TTA AGA CTT ATT GAA 2907
Ala His Phe Leu Ala Lys Ser Leu Asp Asp Ala Leu Arg Leu Ile Glu
90 95 100

CAA CCG GAA TTG GCA AGT AAA GTA GAC ATG GTT TGG ATA GTC GGA GGC 2955
Gln Pro Glu Leu Ala Ser Lys Val Asp Met Val Trp Ile Val Gly Gly
105 110 115

AGT TCT GTT TAC CAG GAA GCC ATG AAT CAA CCA GGC CAC CTC AGA CTC 3003
Ser Ser Val Tyr Gln Glu Ala Met Asn Gln Pro Gly His Leu Arg Leu
120 125 130

TTT GTG ACA AGG ATC ATG CAG GAA TTT GAA AGT GAC ACG TTT TTC CCA 3051
Phe Val Thr Arg Ile Met Gln Glu Phe Glu Ser Asp Thr Phe Phe Pro
135 140 145 150

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Thr	Ala	Pro	Gly	Ala	Ala	Gly	Arg	Cys	Pro	Pro	Thr	Thr	Trp	Leu
1				5					10					15	
Gly	Ser	Leu	Leu	Leu	Leu	Val	Cys	Leu	Leu	Ala	Ser	Arg	Ser	Ile	Thr
			20					25					30		
Glu	Glu	Val	Ser	Glu	Tyr	Cys	Ser	His	Met	Ile	Gly	Ser	Gly	His	Leu
		35					40					45			
Gln	Ser	Leu	Gln	Arg	Leu	Ile	Asp	Ser	Gln	Met	Glu	Thr	Ser	Cys	Gln
	50					55					60				
Ile	Thr	Phe	Glu	Phe	Val	Asp	Gln	Glu	Gln	Leu	Lys	Asp	Pro	Val	Cys
65					70					75					80
Tyr	Leu	Lys	Lys	Ala	Phe	Leu	Leu	Val	Gln	Asp	Ile	Met	Glu	Asp	Thr
				85					90					95	
Met	Arg	Phe	Arg	Asp	Asn	Thr	Pro	Asn	Ala	Ile	Ala	Ile	Val	Gln	Leu
			100					105					110		
Gln	Glu	Leu	Ser	Leu	Arg	Leu	Lys	Ser	Cys	Phe	Thr	Lys	Asp	Tyr	Glu
		115					120					125			
Glu	His	Asp	Lys	Ala	Cys	Val	Arg	Thr	Phe	Tyr	Glu	Thr	Pro	Leu	Gln
	130					135					140				
Leu	Leu	Glu	Lys	Val	Lys	Asn	Val	Phe	Asn	Glu	Thr	Lys	Asn	Leu	Leu
145					150					155					160
Asp	Lys	Asp	Trp	Asn	Ile	Phe	Ser	Lys	Asn	Cys	Asn	Asn	Ser	Phe	Ala
				165					170					175	
Glu	Cys	Ser	Ser	Gln	Asp	Val	Val	Thr	Lys	Pro	Asp	Cys	Asn	Cys	Leu
			180					185					190		
Tyr	Pro	Lys	Ala	Ile	Pro	Ser	Ser	Asp	Pro	Ala	Ser	Val	Ser	Pro	His
		195					200					205			
Gln	Pro	Leu	Ala	Pro	Ser	Met	Ala	Pro	Val	Ala	Gly	Leu	Thr	Trp	Glu
	210					215					220				
Asp	Ser	Glu	Gly	Thr	Glu	Gly	Ser	Ser	Leu	Leu	Pro	Gly	Glu	Gln	Pro
225					230					235					240
Leu	His	Thr	Val	Asp	Pro	Gly	Ser	Ala	Lys	Gln	Arg	Pro	Pro	Arg	
				245					250					255	

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 187 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly
1 5 10 15

Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
20 25 30

Lys Tyr Phe Gln Arg Met Thr Thr Thr Ser Ser Val Glu Gly Lys Gln
35 40 45

Asn Leu Val Ile Met Gly Arg Lys Thr Trp Phe Ser Ile Pro Glu Lys
50 55 60

Asn Arg Pro Leu Lys Asp Arg Ile Asn Ile Val Leu Ser Arg Glu Leu
65 70 75 80

Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Asp
85 90 95

Ala Leu Arg Leu Ile Glu Gln Pro Glu Leu Ala Ser Lys Val Asp Met
100 105 110

Val Trp Ile Val Gly Gly Ser Ser Val Tyr Gln Glu Ala Met Asn Gln
115 120 125

Pro Gly His Leu Arg Leu Phe Val Thr Arg Ile Met Gln Glu Phe Glu
130 135 140

Ser Asp Thr Phe Phe Pro Glu Ile Asp Leu Gly Lys Tyr Lys Leu Leu
145 150 155 160

Pro Glu Tyr Pro Gly Val Leu Ser Glu Val Gln Glu Glu Lys Gly Ile
165 170 175

Lys Tyr Lys Phe Glu Val Tyr Glu Lys Lys Asp
180 185

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 553 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..32
(D) OTHER INFORMATION: /label= leader_peptide

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 33..553
(D) OTHER INFORMATION: /note= "Mature MCSF peptide"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Ala Pro Gly Ala Ala Gly Arg Cys Pro Pro Thr Thr Trp Leu
1   5   10   15

Gly Ser Leu Leu Leu Leu Val Cys Leu Leu Ala Ser Arg Ser Ile Thr
            20   25   30

Glu Glu Val Ser Glu Tyr Cys Ser His Met Ile Gly Ser Gly His Leu
        35   40   45

Gln Ser Leu Gln Arg Leu Ile Asp Ser Gln Met Glu Thr Ser Cys Gln
    50   55   60

Ile Thr Phe Glu Phe Val Asp Gln Glu Gln Leu Lys Asp Pro Val Cys
65   70   75   80

Tyr Leu Lys Lys Ala Phe Leu Leu Val Gln Asp Ile Met Glu Asp Thr
                85   90   95

Met Arg Phe Arg Asp Asn Thr Pro Asn Ala Ile Ala Ile Val Gln Leu
            100   105   110

Gln Glu Leu Ser Leu Arg Leu Lys Ser Cys Phe Thr Lys Asp Tyr Glu
        115   120   125

Glu His Asp Lys Ala Cys Val Arg Thr Phe Tyr Glu Thr Pro Leu Gln
    130   135   140

Leu Leu Glu Lys Val Lys Asn Val Phe Asn Glu Thr Lys Asn Leu Leu
145   150   155   160

Asp Lys Asp Trp Asn Ile Phe Ser Lys Asn Cys Asn Asn Ser Phe Ala
                165   170   175

Glu Cys Ser Ser Gln Asp Val Val Thr Lys Pro Asp Cys Asn Cys Leu
            180   185   190

Tyr Pro Lys Ala Ile Pro Ser Ser Asp Pro Ala Ser Val Ser Pro His
        195   200   205

Gln Pro Leu Ala Pro Ser Met Ala Pro Val Ala Gly Leu Thr Trp Glu
    210   215   220

Asp Ser Glu Gly Thr Glu Gly Ser Ser Leu Leu Pro Gly Glu Gln Pro
225   230   235   240

Leu His Thr Val Asp Pro Gly Ser Ala Lys Gln Arg Pro Pro Arg Ser
                245   250   255

Thr Cys Gln Ser Phe Glu Pro Pro Glu Thr Pro Val Val Lys Asp Ser
            260   265   270

Thr Ile Gly Gly Ser Pro Gln Pro Arg Pro Ser Val Gly Ala Phe Asn
        275   280   285

Pro Gly Met Glu Asp Ile Leu Asp Ser Ala Met Gly Thr Asn Trp Val
    290   295   300

Pro Glu Glu Ala Ser Gly Glu Ala Ser Glu Ile Pro Val Pro Gln Gly
305   310   315   320

Thr Glu Leu Ser Pro Ser Arg Pro Gly Gly Gly Ser Met Gln Thr Glu
325 330 335

Pro Ala Arg Pro Ser Asn Phe Leu Ser Ala Ser Ser Pro Leu Pro Ala
340 345 350

Ser Ala Lys Gly Gln Gln Pro Ala Asp Val Thr Gly Thr Ala Leu Pro
355 360 365

Arg Val Gly Pro Val Arg Pro Thr Gly Gln Asp Trp Asn His Thr Pro
370 375 380

Gln Lys Thr Asp His Pro Ser Ala Leu Leu Arg Pro Pro Glu Pro Gly
385 390 395 400

Ser Pro Arg Ile Ser Ser Leu Arg Pro Gln Gly Leu Ser Asn Pro Ser
405 410 415

Thr Leu Ser Ala Gln Pro Gln Leu Ser Arg Ser His Ser Ser Gly Ser
420 425 430

Val Leu Pro Leu Gly Glu Leu Glu Gly Arg Arg Ser Thr Arg Asp Arg
435 440 445

Arg Ser Pro Ala Glu Pro Glu Gly Gly Pro Ala Ser Glu Gly Ala Ala
450 455 460

Arg Pro Leu Pro Arg Phe Asn Ser Val Pro Leu Thr Asp Thr Gly His
465 470 475 480

Glu Arg Gln Ser Glu Gly Ser Ser Ser Pro Gln Leu Gln Glu Ser Val
485 490 495

Phe His Leu Leu Val Pro Ser Val Ile Leu Val Leu Leu Ala Val Gly
500 505 510

Gly Leu Leu Phe Tyr Arg Trp Arg Arg Arg Ser His Gln Glu Pro Gln
515 520 525

Arg Ala Asp Ser Pro Leu Glu Gln Pro Glu Gly Ser Pro Leu Thr Gln
530 535 540

Asp Asp Arg Gln Val Glu Leu Pro Val
545 550

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 10, line 5 of the description

A. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet ☐

Name of depositary institution

America Type Culture Collection

Address of depositary institution (including postal code and country)

12301 Parklawn Drive
Rockville, MD 20852 United States of America

Date of deposit	Accession Number
December 11, 1992	ATCC 75378

B. ADDITIONAL INDICATIONS (leave blank if not applicable). This information is continued on a separate attached sheet ☐

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

D. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)

(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3 AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

To: (Name and Address of Depositor or Attorney)

Genetics Institute, Inc.
Attention: Andrew J. Dorner
87 CambridgePark Drive
Cambridge, MA 02140

Deposited on Behalf of: Genetics Institute, Inc.

Identification Reference by Depositor: ATCC Designation

DNA Plasmid, p MEMCMat 75378

The deposit was accompanied by: __ a scientific description __ a proposed taxonomic description indicated above.

The deposit was received December 11, 1992 by this International Depository Authority and has been accepted.

AT YOUR REQUEST:

X We will inform you of requests for the strain for 30 years.

T strain will be made available if a patent office signatory to the Budapest Treaty certifies one's rignt to receive, or if a U.S. Patent is issued citing the strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years after the date of deposit, and for a period of at least five years after the most recent request for a sample. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested December 21, 1992. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Rockville, Md. 20852 USA

Signature of person having authority to represent ATCC:

Bobbie A. Brandon Date: January 4, 1993
Bobbie A. Brandon, Head, ATCC Patent Depository

cc: Steven R. Lazar

Form BP4/9

International Application No: PCT/

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 13, line 5-6 of the description [1]

A. IDENTIFICATION OF DEPOSIT [1]

Further deposits are identified on an additional sheet ☐ [3]

Name of depositary institution [4]

America Type Culture Collection

Address of depositary institution (including postal code and country) [4]

12301 Parklawn Drive
Rockville, MD 20852 United States of America

Date of deposit [5]	Accession Number [6]
Eebruary 19, 1993	ATCC CRL 11275

B. ADDITIONAL INDICATIONS [7] (leave blank if not applicable). This information is continued on a separate attached sheet ☐

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE [8] (if the indications are not for all designated States)

D. SEPARATE FURNISHING OF INDICATIONS [9] (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later [9] (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☐ This sheet was received with the international application when filed (to be checked by the receiving Office)

Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau [10]

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

American Type Culture Collection

12301 Parklawn Drive • Rockville, MD 20852 USA • Telephone: (301)231-5520 Telex: 898-055 ATCCNORTH • FAX: 301-770-2587

Create ATCC CRL 11275 Folder

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3 AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

To: (Name and Address of Depositor or Attorney)

Genetics Institute
Attention: S. Robert Adamson
87 CambridgePark Drive
Cambridge, MA 02140

Deposited on Behalf of: Genetics Institute

Identification Reference by Depositor: ATCC Designation

Chinese Hamster Ovary Cells, (7/1)CHO - H1 - 2.0 CRL 11275

The deposit was accompanied by: __ a scientific description __ a proposed taxonomic description indicated above.

The deposit was received February 19, 1993 by this International Depository Authority and has been accepted.

AT YOUR REQUEST:

X We will inform you of requests for the strain for 30 years.

The strain will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years after the date of deposit, and for a period of at least five years after the most recent request for a sample. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested February 24, 1993. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Rockville, Md. 20852 USA

Signature of person having authority to represent ATCC:

Bobbie A. Brandon Date: February 25, 1993
Bobbie A. Brandon, Head, ATCC Patent Depository

cc: Steven R. Lazar

Form BP4/9

SUBSTITUTE SHEET (RULE 26)

We claim:

1. An isolated and purified DNA molecule consisting essentially of the DNA sequence of Sequence ID No. 1.

2. An isolated and purified DNA molecule according to claim 1, wherein the DNA molecule encodes the production of a 255 amino acid polypeptide, comprising a 32 amino acid leader sequence from M-CSF and a 223 amino acid polypeptide having M-CSF activity, said polypeptide lacking the carboxy-terminal propeptide sequence present in wild-type recombinant M-CSF polypeptide before intracellular processing when produced by mammalian cells.

3. An isolated and purified DNA molecule comprising the plasmid pMemcMAT.

4. An isolated and purified DNA molecule consisting essentially of a DNA sequence encoding the polypeptide comprised of the amino acid sequence of Sequence ID No. 2.

5. An isolated and purified DNA molecule according to claim 4, wherein the DNA molecule encodes the production of a 255 amino acid polypeptide, comprising a 32 amino acid leader sequence from M-CSF and a 223 amino acid polypeptide having M-CSF activity, said polypeptide lacking the carboxy-terminal propeptide sequence present in wild-type recombinant M-CSF polypeptide before intracellular processing when produced by mammalian cells.

6. A method of producing MCSF-223 protein comprising:

a) culturing in a suitable culture medium cells transformed with a DNA molecule according to claim 1; and

b) isolating and purifying said protein from said culture medium.

7. A method of producing MCSF-223 protein comprising:

a) culturing in a suitable culture medium cells transformed with the DNA molecule according to claim 3;

b) isolating and purifying said protein from said culture medium.

8. A mammalian cell line capable of producing MCSF-223 protein, said mammalian cell line comprising a DNA sequence according to claim 1.

9. A mammalian cell line capable of producing MCSF-223 protein, said mammalian cell line comprising a DNA sequence according to claim 2.

10. A method of purifying MCSF-223 protein comprising the following steps in sequence:

a) ultrafiltration/diafiltration;

b) QAE- Sephrose chromatography;

c) ceramic hydroxyapatite chromatography;

d) Phenyl-Toyoperl chromatography; and

e) diafiltration.

FIGURE 1

MTAPGAAGRCPPTTWLGSLLLLVCLLASRSIT -32

E E V S E Y C S H M I G S G H L Q S L Q R L I D S 25
Q M E T S C Q I T F E F V D Q E Q L K D P V C Y L 50
K K A F L L V Q D I M E D T M R F R D N T P N A I 75
A I V Q L Q E L S L R L K S C F T K D Y E E H D K 100
A C V R T F Y E T P L Q L L E K V K N V F N E T K 125
N L L D K D W N I F S K N C N N S F A E C S S Q D 150
V V T K P D C N C L Y P K A I P S S D P A S V S P 175
H Q P L A P S M A P V A G L T W E D S E G T E G S 200
S L L P G E Q P L H T V D P G S A K Q R P P R **^ 223

S T C Q S F E P P E T P V V K D S T I G G S P Q P 248
R P S V G A F N P G M E D I L D S A M G T N W V P 273
E E A S G E A S E I P V P Q G T E L S P S R P G G 298
G S M Q T E P A R P S N F L S A S S P L P A S A K 323
G Q Q P A D V T G T A L P R V G P V R P T G Q D W 348
N H T P Q K T D H P S A L L R O P P E P G S P R I 373
S S L R P Q G L S N P S T L S A Q P Q L S R S H S 398
S G S V L P L G E L E G R R S T R D R R S P A E P 423
E G G P A S E G A A R P L P R F N S V P L T D T G 448
H E R Q S E G S S S P Q L Q E S V F H L L V P S V 473
I L V L L A V G G L L F Y R W R R R S H Q E P Q R 498
A D S P L E Q P E G S P L T Q D D R Q V E L P V 522

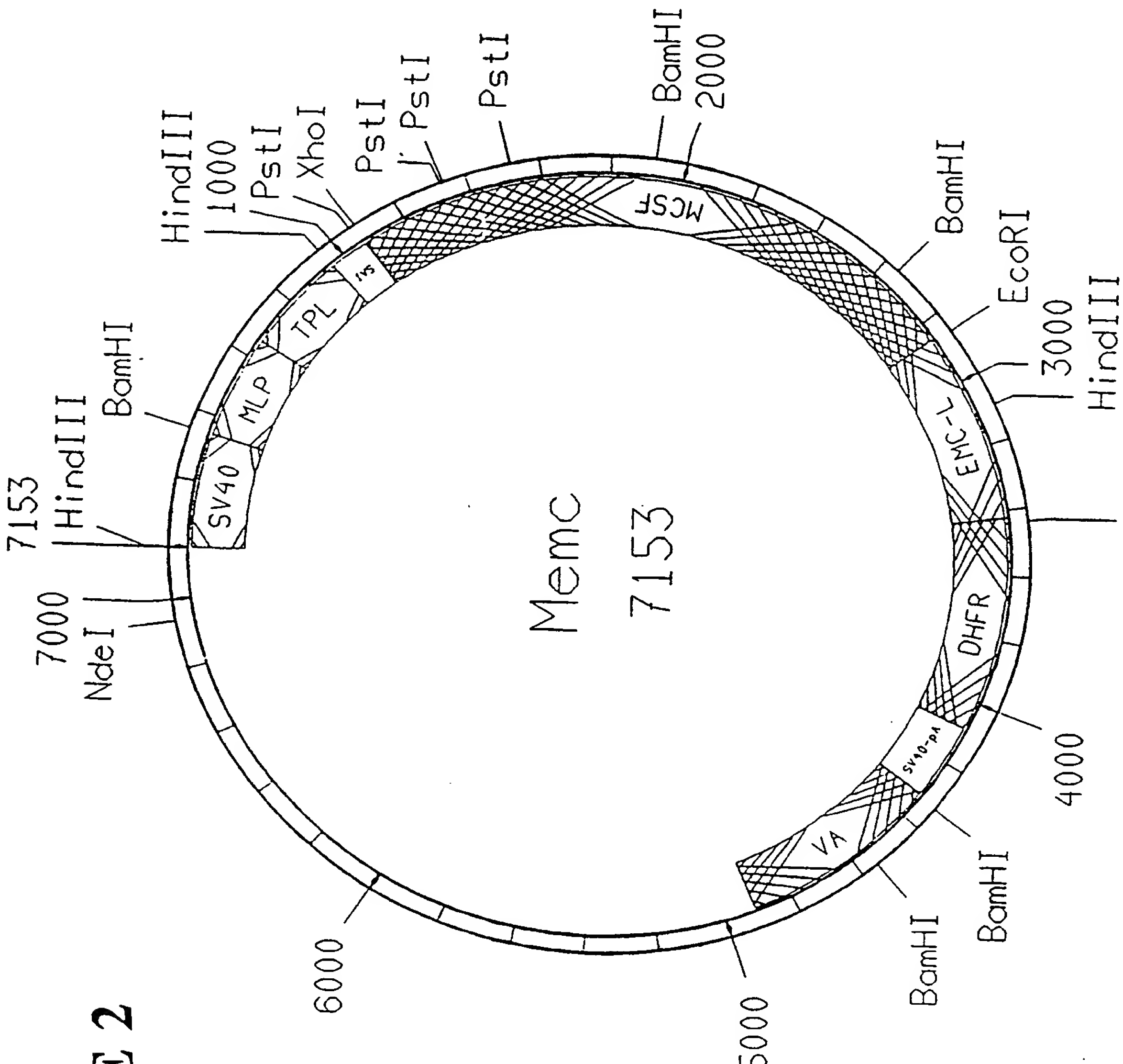
Memc
7153
MCSF
IVS
TPL
MLP
SV40
EMC-L
DHFR
SV40-pA
VA
HindIII
1000
PstI
XhoI
PstI
PstI
PstI
BamHI
2000
BamHI
EcoRI
3000
HindIII
HindIII
XhoI
4000
BamHI
BamHI
5000
6000
7000
NdeI
7153
HindIII
HindIII
BamHI

FIGURE 2

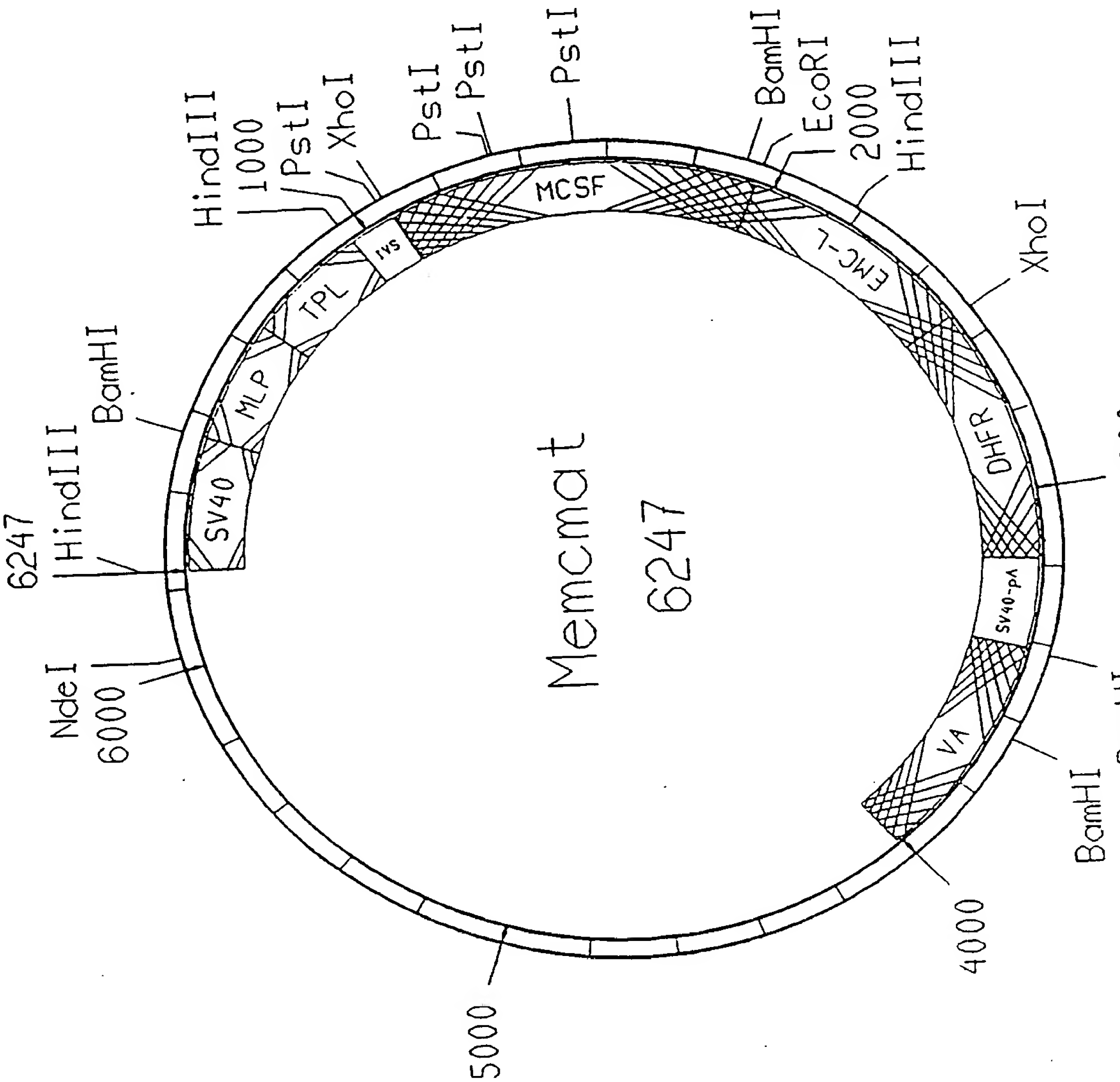
Memcmat
6247
MCSF
IVS
TPL
MLP
SV40
EMC-L
DHFR
SV40-pA
VA
PstI
PstI
PstI
PstI
BamHI
EcoRI
2000
HindIII
XhoI
3000
BamHI
BamHI
4000
5000
NdeI
6000
6247
HindIII
HindIII
BamHI
HindIII
1000
XhoI

FIGURE 3